American Journal of Epidemiology
Published by Oxford University Press on behalf of the Johns Hopkins Bloomberg School of Public Health 2018.

Vol. 188, No. 3
DOI: 10.1093/aje/kwy248
Advance Access publication:
December 18, 2018

Practice of Epidemiology

Causal Mediation Analysis With Observational Data: Considerations and Illustration Examining Mechanisms Linking Neighborhood Poverty to Adolescent Substance Use

Kara E. Rudolph*, Dana E. Goin, Diana Paksarian, Rebecca Crowder, Kathleen R. Merikangas, and Elizabeth A. Stuart

* Correspondence to Dr. Kara E. Rudolph, Department of Emergency Medicine, University of California, Davis, 2315 Stockton Boulevard, Sacramento, CA 95817 (e-mail: kerudolph@ucdavis.edu).

Initially submitted February 14, 2018; accepted for publication October 26, 2018.

Understanding the mediation mechanisms by which an exposure or intervention affects an outcome can provide a look into what has been called a "black box" of many epidemiologic associations, thereby providing further evidence of a relationship and possible points of intervention. Rapid methodologic developments in mediation analyses mean that there are a growing number of approaches for researchers to consider, each with its own set of assumptions, advantages, and disadvantages. This has understandably resulted in some confusion among applied researchers. Here, we provide a brief overview of the mediation methods available and discuss points for consideration when choosing a method. We provide an in-depth explication of 2 of the many potential estimators for illustrative purposes: the Baron and Kenny mediation approach, because it is the most commonly used, and a recently developed approach for estimating stochastic direct and indirect effects, because it relies on far fewer assumptions. We illustrate the decision process and analytical procedure by estimating potential school- and peer-based mechanisms linking neighborhood poverty to adolescent substance use in the National Comorbidity Survey Adolescent Supplement.

adolescent; mediation; neighborhood; stochastic intervention; substance use

Abbreviation: TMLE, targeted minimum loss-based estimation.

Rapid methodologic developments in mediation analyses mean that there are a growing number of approaches for researchers to consider, each with its own set of assumptions, advantages, and disadvantages. This has understandably resulted in some confusion among applied researchers. Here, we provide a brief overview of the mediation methods available (building on previous work (1–10)), discuss points for consideration when choosing a method, and illustrate the decision process and analytical procedure by estimating potential school- and peer-based mechanisms linking neighborhood poverty to adolescent substance use in the National Comorbidity Survey Replication Adolescent Supplement (11).

We consider 3 general types of path-specific, causal mediation estimands (i.e., types of effects) (12): 1) controlled direct effects (13), 2) natural direct and indirect effects (13), and 3) stochastic (also called randomized interventional) direct and indirect effects (14–16). We define and discuss each of these mathematically and intuitively. We then address the first step of the decision process: choosing the estimand that best reflects the research question and whose identifying assumptions are plausible given our knowledge of the structural causal model (17) (which might be conveyed by a directed acyclic graph) representing the research question. We then discuss the second step of the process: choosing an estimator, again based on assumptions that we believe to be reasonable. We provide an in-depth explication of 2 of the many potential estimators for illustrative purposes, taking a case-study approach. We illustrate: 1) the Baron and Kenny mediation approach (18, 19), because it is the most commonly used in the epidemiologic literature; and 2) a recently developed approach for estimating stochastic direct and indirect effects (16), because it both represents a contrast with the Baron and Kenny approach by relying on fewer assumptions and was used to estimate mediated effects in a similar application to the one we consider in the applied portion of

this work. We show results from our illustrative example for the 2 approaches contrasted, discuss their differences, and compare the assumptions and implications.

ILLUSTRATIVE RESEARCH QUESTION AND DATA SET

Several studies have linked living in poor neighborhoods to risk of problematic drug and alcohol use among adolescents (20–25), the public health importance of which is described in Web Appendix 1 (available at https://academic.oup.com/aje). Identifying mechanisms that catalyze these effects might lend credence to the associations and present additional specific targets for intervention. There is some evidence of the school and peer environments mediating the relationship between neighborhood and adolescent substance use (25–27). However, this prior evidence was limited by analytical approaches that required overly restrictive assumptions. Recent work relaxing these assumptions found weak evidence of mediation by aspects of the peer environment and no mediation by aspects of the school environment (28). However, that work used Section 8 housing voucher receipt as a surrogate for neighborhood disadvantage, focused on a select number of US cities, and did not have diagnoses of substance use disorder. In our illustrative example, we examined whether these previous findings (28) generalize to a representative sample of urban US adolescents using a composite measure of neighborhood disadvantage and a *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition*, diagnosis of substance use disorder in addition to drug, alcohol, and tobacco use outcomes. A directed acyclic graph reflecting this research question is shown in Figure 1.

We examined this research question in the National Comorbidity Survey Replication Adolescent Supplement, a nationally representative survey of US adolescents conducted during 2001–2004. Additional details regarding the illustrative example are given in Web Appendix 2. Details of the sampling design and procedures have been published previously (11, 29, 30). Details of our analytical sample are provided in Web Appendix 2.1 and in Web Figure 1. Written informed consent was provided by parents and assent by adolescents. Study procedures were approved by the human subjects committees of Harvard Medical School and the University of Michigan. Our exposure of interest was neighborhood disadvantage. We considered 4 binary mediators related to the school and peer environments: 1) high rates of school violent crime, 2) school security presence, 3) whether most or all of the adolescent's friends and siblings ever use marijuana or other drugs, and 4) the adolescent never having participated in an after-school sport or club. We considered 6 binary substance use outcomes: 1) lifetime cigarette use, 2) lifetime alcohol use, 3) problematic alcohol use, 4) lifetime marijuana use, 5) problematic drug use, and 6) past-year *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition*, diagnosis of substance use abuse or dependence. Measurement details of each of these variables and covariates are described in Web Appendix 2.2.

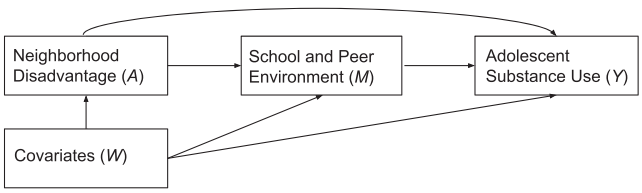

Figure 1. Directed acyclic graph of a structural causal model of mediation of the relationship between neighborhood disadvantage and adolescent substance use by the school and peer environments.

OVERVIEW OF MEDIATION ESTIMANDS AND ESTIMATORS

There are 3 main types of path-specific, causal mediation estimands that represent the direct effect of the exposure on the outcome, not operating through the mediator, and the indirect effect of the exposure on the outcome that operates through the mediator. Controlled direct effects have been termed "prescriptive," because they hypothesize intervening directly on the mediator, assigning the same value to everyone (13). In contrast, natural direct and indirect effects have been termed "descriptive," because they hypothesize assigning mediator values based on counterfactual values associated with the exposure scenario of interest and are thus used to "describe" the mechanism of mediation (13). Stochastic direct and indirect effects are also "descriptive" in this sense, but rather than assigning individuals their own counterfactual values of the mediator, values are drawn from the distribution that corresponds to the exposure scenario of interest and strata of covariates (10, 15, 16, 31). These are each described in more detail below and in Table 1.

Notation

We first define notation. We observe data $O = (W, A, M, Y)$, where W represents covariates, A represents exposure (e.g., neighborhood disadvantage), M represents the mediator (e.g., school and peer environment), and Y represents the outcome (e.g., adolescent substance use). Under this notation, the direct effect is the effect $A \rightarrow Y$, not through M, and the indirect effect is the effect $A \rightarrow M \rightarrow Y$.

Each causal mediation effect represents a contrast of potential outcomes (32) that are a function of both A and M. Potential outcomes represent counterfactual or "what-if" scenarios and can be thought of as what would have happened under alternative histories. We observe only one of these potential outcomes in reality. For example, $Y_{a,m}$ represents the potential outcome setting A to value a and M to value m, possibly contrary to fact. We can also have potential mediator values that are a function of A. For example, M_a represents the potential mediator setting A to a. We make consistency and positivity assumptions throughout. For consistency, this means that the counterfactual quantity $Y_{a,m}$ equals the observed value of Y when $A = a$ and $M = m$. For positivity, this means that there is a nonzero probability that M is equal to each of its potential values conditional on A and W and a nonzero probability that $A = a$ or $A = a^*$ conditional on W.

Controlled direct effects

Controlled direct effects are defined as: $E(Y_{a,m}) - E(Y_{a^*,m})$ (13). In terms of the illustrative example, the controlled direct

Table 1. Mediation Estimand Definitions, Descriptions, and Assumptions

Estimand	Description	Identifying Assumptions in Addition to Positivity and Consistency
Controlled direct effect $E(Y_{a,m}) - E(Y_{a^*,m})$	Difference in the expected value of Y setting A to a versus a^* and in both cases setting M to m	1. No unmeasured confounding between A and Y ($A \perp Y_{a,m} \mid W$). 2. No unmeasured confounding between M and Y ($M \perp Y_{a,m} \mid W, A$).
Natural direct effect $E(Y_{a,M_{a^*}}) - E(Y_{a^*,M_{a^*}})$ Natural indirect effect $E(Y_{a,M_a}) - E(Y_{a,M_{a^*}})$	Difference in the expected value of Y setting A to a versus a^* and in both cases letting M be the value that it would naturally be under a^* Difference in the expected value of Y in both cases setting A to a and contrasting M under a versus a^*	1. No unmeasured confounding between A and Y ($A \perp Y_{a,m} \mid W$). 2. No unmeasured confounding between M and Y ($M \perp Y_{a,m} \mid W, A$). 3. No unmeasured confounding of $A - M$ ($A \perp M_a \mid W$). 4. No measured or unmeasured posttreatment confounding of the $M - Y$ relationship ($M_{a^*} \perp Y_{a,m} \mid W$). 5. Y_a is equivalent to Y_{a,M_a}.
Stochastic direct effect $E(Y_{a,g_{M\mid a^*,W}}) - E(Y_{a^*,g_{M\mid a^*,W}})$ Stochastic indirect effect $E(Y_{a,g_{M\mid a,W}}) - E(Y_{a,g_{M\mid a^*,W}})$	Difference in the population average of Y setting A to a versus a^* and in both cases drawing the value of M from a distribution of M conditional on $A = a^*$ and the individual's set of covariate values, W Difference in the population average of Y in both cases setting A to a and contrasting drawing the value of M from a distribution of M conditional on $A = a$ versus $A = a^*$ and the individual's set of covariate values, W	1. No unmeasured confounding between A and Y ($A \perp Y_{a,m} \mid W$). 2. No unmeasured confounding between M and Y ($M \perp Y_{a,m} \mid W, A$). 3. No unmeasured confounding of $A - M$ ($A \perp M_a \mid W$).

Abbreviations: A, treatment; M, mediator; W, covariates; Y, outcome.

effect would be the population average of the difference in potential outcomes contrasting living in a disadvantaged versus nondisadvantaged neighborhood and setting peer substance use to the same level for everyone.

These effects are identified if there is no unmeasured confounding of $A - Y$ or $M - Y$ in addition to positivity and consistency assumptions (Table 1). If one is comfortable with these identifying assumptions, one then needs to think about whether the estimand aligns with the research question. For example, controlled direct effects might make sense in thinking about medical providers setting the dosage of a prescription to a standard amount ($M = m$) (2). However, there is no indirect effect counterpart, as it is strictly defined, because there is no contrast of controlled potential outcomes, like $E(Y_{a,m} - Y_{a,m^*})$, that allows A to affect M (12, 13). Instead, contrasts like $E(Y_{a,m} - Y_{a,m^*})$ represent the effect $M \rightarrow Y$ and have been termed controlled mediator effects or controlled direct effects of the mediator (33, 34). In addition, if there is interaction between the exposure and the mediator in the true outcome model, the effect estimate will differ based on the value of the mediator chosen (6).

Natural direct and indirect effects

The natural direct effect is defined as $E(Y_{a,M_{a^*}}) - E(Y_{a^*M_{a^*}})$ and natural indirect effect is defined as $E(Y_{a,M_a}) - E(Y_{a,M_{a^*}})$, where M_a and M_{a^*} are individual-specific, counterfactual values of the mediator had A been set to a or a^*, respectively (13). In terms of the illustrative example, the natural direct effect is the population average of the difference in the individual-specific potential outcomes contrasting if the individual lived in a disadvantaged versus nondisadvantaged neighborhood and in both cases letting their level of peer substance use be at their potential level in the nondisadvantaged neighborhood.

Natural direct effects can also be written as the weighted average of controlled direct effects at each level of $M = m$, $\sum_m \{E(Y_{a,m}) - E(Y_{a^*,m})\} P(M_{a^*} = m)$ (2). Thus, one can see that, in the absence of interaction between A and M on Y, the controlled and natural direct effects will be equivalent (2, 35).

These effects are identified if there are no unmeasured confounders of 1) $A - Y$, 2) $M - Y$, and 3) $A - M$ and no measured or unmeasured posttreatment $M - Y$ confounders in addition to positivity and consistency assumptions (Table 1). Thus, these effects have 2 assumptions more than their controlled counterparts. This last identification assumption, $M_{a^*} \perp Y_{a,m} \mid W$, is a "cross-world" independence assumption, because it simultaneously assumes a world in which $A = a$ and another in which $A = a^*$, so cannot be tested in reality. If one is comfortable with the identifying assumptions, one then needs to think about whether the estimand makes sense in terms of the research question. These estimands make sense if the research question involves intervening directly on A but not on M (13). One example is the research question we examine here, where it might make sense to intervene on neighborhood exposure (through a housing intervention, for example) and where we are curious about how the possible downstream consequences of such an intervention (e.g., in terms of the school and peer environments) might act to affect adolescent health. Natural direct and indirect effects also have the advantage of adding to the total effect.

However, assuming no measured or unmeasured posttreatment $M - Y$ confounding might be frequently violated in practice. For example, it is violated whenever treatment assignment A is hypothesized to act through adherence to

the treatment, Z, because adherence might affect both the mediator and outcome and would of course be affected by treatment assignment (e.g., in any instrumental variable mediation scenario) (16). In addition, it would also be violated in longitudinal data structures where time-varying confounders would be affected by treatment and in turn affect both the mediator and outcome (15). If this assumption seems problematic for the research question, or if the "cross-world" component is viewed as problematic, then the researcher might want to consider stochastic direct and indirect effects or controlled direct effects, depending on which is more closely aligned with the research question.

Stochastic direct and indirect effects

There are 2 versions of stochastic direct and indirect effects, one that conditions on a posttreatment $M - Y$ confounder, Z, and one that marginalizes over Z. The conditional stochastic direct effect is defined as $E(Y_{a,g_{M|Z,a^*,W}}) - E(Y_{a^*,g_{M|Z,a^*,W}})$, and the indirect effect is $E(Y_{a,g_{M|Z,a,W}}) - E(Y_{a,g_{M|Z,a^*,W}})$, where $g_{M|Z,a^*,W} = P(M{=}1|Z, a^*, W)$ represents a stochastic draw from the distribution of M conditional on Z, a^*, and W. The marginal versions are defined similarly: $E(Y_{a,g_{M|a^*,W}}) - E(Y_{a^*,g_{M|a^*,W}})$ and $E(Y_{a,g_{M|a,W}}) - E(Y_{a,g_{M|a^*,W}})$ for the direct and indirect effects, respectively, where $g_{M|a^*,W} = \sum_z P(M{=}1|Z = z, W)P(Z = z|A = a^*, W)$. In terms of the illustrative example, the marginal stochastic direct effect is the population average of the difference in individual-specific potential outcomes contrasting if the individual had lived in a disadvantaged versus nondisadvantaged neighborhood and in both cases letting their level of peer substance use be drawn from a distribution of peer substance use in nondisadvantaged neighborhoods, conditional on covariates.

Stochastic direct and indirect effects are identified if there are no unmeasured confounders of 1) $A - Y$, 2) $M - Y$, and 3) $A - M$ in addition to positivity and consistency assumptions (Table 1). Thus, these effects have 1 assumption more than their controlled counterparts by assuming no unmeasured confounding of $A - M$ but have fewer assumptions than their natural counterparts by allowing for measured posttreatment $M - Y$ confounders and avoiding a cross-world assumption. In the absence of a posttreatment $M - Y$ confounder, as is the case in the scenario we consider here (Figure 1), these effects are analogous to the population-average natural direct and indirect effects (10, 15), with only slight differences in point estimates as discussed in VanderWeele and Tchetgen Tchetgen (15) in addition to the difference in their interpretation. This makes intuitive sense, because—in this scenario—taking the average of individual stochastic draws from a distribution of mediator values (as for stochastic direct/indirect effects) coincides with taking the average across potential mediator values across individuals (as for natural direct/indirect effects), making the difference largely semantic in interpretation.

It is not clear whether or not there are substantive reasons to choose stochastic as opposed to natural mediation effects. Others have speculated that for research questions where A is considered a fixed characteristic of an individual, such as race, it might be easier to imagine contrasting distributions of potential mediator values as opposed to contrasting individual-specific potential mediator values (15). However, if A is not manipulable, it likely does not make sense to think of either individual-level potential outcome values or individual-level potential mediator values.

If the researcher chooses to estimate stochastic mediation effects, he or she also needs to decide whether to use the estimands that condition on Z versus marginalize over Z. There might be substantive and/or practical reasons to choose one over the other. In terms of the indirect effect, conditioning on Z estimates the pathway $A \rightarrow M \rightarrow Y$, not through Z. Marginalizing over Z estimates the combined pathways: $A \rightarrow M \rightarrow Y$ and $A \rightarrow Z \rightarrow M \rightarrow Y$, thus allowing the effect of A to work through Z.

For example, one would want to condition on Z in settings where mediator values make sense only in the presence or absence of Z, such as when Z represents survival, because then one is drawing the mediator value from a distribution that includes only other survivors (36). In contrast, one would want to marginalize over Z in instrumental variable settings where instrument A affects M and Y only through its effect on Z (16). In such settings, conditional stochastic indirect effects would necessarily be zero as would indirect effects from approaches like sequential mediation analysis (37), because there is no direct effect of A on M that does not go through Z. In the absence of substantive reasons to choose one over the other, a practical reason to marginalize over Z is that for most estimation approaches, one would not need to specify a model for Z.

Estimation

Choosing an estimand is the first step in the mediation analysis process. Once an estimand has been deemed appropriate (based on how closely it reflects the research question and how reasonable its identifying assumptions are to hold), the researcher next chooses an estimator. Table 2 provides citations for different estimators according to estimand type. This table is not meant to be comprehensive but instead to serve as a starting point for researchers to explore estimation possibilities. We describe and illustrate 2 of these methods in additional detail below: the Baron and Kenny approach and targeted minimum loss-based estimation (TMLE) for stochastic marginal direct and indirect effects (16, 18, 19).

Baron and Kenny parametric regression approach. The Baron and Kenny approach (18, 19) is widely used across public health and the social sciences, so it is one of the approaches we illustrate here. One reason for its popularity might be its simplicity of implementation. It involves 2 regressions: 1) $E(M|a, w) = \beta_0 + \beta_1 a + \beta_2 w$, and 2) $E(Y|m, a, w) = \theta_0 + \theta_1 a + \theta_2 w + \theta_3 m$ (19). It makes an additional assumption of no interaction of A and M on Y, which means that the controlled direct effect is equivalent to the natural direct effect. The natural or controlled direct effect conditional on W at a fixed level of m equals θ_1. The natural indirect effect equals $\beta_1 \times \theta_3$ with variance $\beta_1^2 \times var(\theta_3) + \theta_3^2 \times var(\beta_1) + var(\beta_1) \times var(\theta_3)$ (3, 19). This variance estimate has appropriate coverage if the estimator for the indirect effect is normally distributed. Products of normal random variables are typically

Table 2. Estimation Strategies by Type of Mediation Estimand

Estimand	Citation of Estimation Method[a]
Controlled direct effects	Baron and Kenny method (18, 19), Robins and Greenland (35), Petersen et al. (2), VanderWeele (55), Goetgeluk et al. (56), Lendle et al. (57)
Natural direct and indirect effects	Baron and Kenny (18, 19), Robins and Greenland (35), Petersen et al. (2), VanderWeele (55), Tchetgen Tchetgen (38), Tchetgen Tchetgen and VanderWeele (58), Taguri and Chiba (59), Vansteelandt and VanderWeele (60), Zheng and van der Laan (39)
Stochastic direct and indirect effects	
Conditional	Zheng and van der Laan (36)
Marginal	VanderWeele and Tchetgen Tchetgen (15), Rudolph et al. (16)

[a] In order of most assumptions required to fewest assumptions required.

skewed, however, so the resulting 95% confidence intervals are likely inaccurate (3).

Although the simplicity of the analysis is appealing, the additional assumptions of 1) no $A - M$ interaction on Y, 2) correctly specified parametric models, and 3) a linear relationship between M and Y (1) might be unrealistic and, together with potentially inaccurate variance estimation, might motivate the researcher to explore other estimation possibilities (of which there are several (2, 15, 16, 36, 38, 39)). For example, in the research question we examined here, it seems possible and perhaps even likely that the degree to which aspects of the school and peer environments influence adolescent substance use might depend on the degree to which the neighborhood is disadvantaged. In addition, assuming correct parametric model specification of 1) how all covariates, W, and neighborhood disadvantage, A, influence aspects of the school and peer environments, M, and 2) how W, A, and M influence substance use, Y, seems unrealistic, given that such social relationships do not rely on known processes. Thus, we might prefer to explore an alternative estimation approach that allows for $A - M$ interactions and reduces reliance on correct parametric model specification.

TMLE semiparametric, data-adaptive approach. The other approach we illustrate, TMLE for marginal stochastic direct and indirect effects (16), achieves the aforementioned goals in that it: 1) estimates a "descriptive" mediation estimand by estimating stochastic direct and indirect effects; 2) allows $A - M$ interactions; 3) does not require the assumption of no posttreatment confounding; 4) reduces reliance on correct parametric model specification by both being doubly robust (meaning that we can get a consistent estimate if either the A and M models are correctly specified or the Y model is correctly specified) and by integrating data-adaptive, machine-learning algorithms into model fitting; and 5) returns theory-based inference that incorporates this data adaptivity (i.e., accurate confidence intervals, calculated from variance estimated as the sample variance of the efficient influence curve (28)). The first step in implementing this estimator is to estimate the stochastic interventions for each mediator, $g_{M|a*,W}$ and $g_{M|a,W}$. We estimate data-dependent versions of these that assume a known, observed distribution of M conditional on W and A. Reasons for this are technical and described elsewhere (16). We then incorporate these estimated stochastic interventions into the TMLE (16). Step-by-step instructions for implementing the TMLE and annotated R code (R Foundation for Statistical Computing, Vienna, Austria) have been published previously (16). Briefly, TMLE is a doubly robust substitution estimator. Using parametric language for simplicity, the estimation strategy models the A, M, and Y distributions and incorporates a "targeting step," which adds robustness against model misspecification. Intuitively, the approach is akin to weighted g-computation, where a parametric outcome model could be weighted by inverse probability weights to add robustness. We provide a more detailed explication in Web Appendix 3 as well as annotated R code in Web Appendix 4. Additionally, we refer the interested reader to the following references to learn more about TMLE in general (40) and about the particular TMLE we employed in estimating stochastic direct and indirect effects (16). Previous research examining a similar research question used this TMLE estimator, so the results from our illustrative example can be compared with this previous work (28).

We note that the Baron and Kenny approach estimates conditional natural direct and indirect effects while the TMLE approach estimates marginal versions of these effects. These approaches will differ if the linear model is misspecified and there is in fact effect modification by $\mathbf{W}$.

ILLUSTRATIVE ANALYSIS AND RESULTS

Statistical approach

For each mediator-outcome combination, we used the Baron and Kenny approach to estimate natural direct and indirect effects, conditional on covariates, and we used TMLE to estimate data-dependent stochastic direct and indirect effects (16). In this case, where we assume no posttreatment $M - Y$ confounder (Figure 1), the stochastic effects are analogous with their natural counterparts, excepting the difference in interpretation. We chose to estimate these types of effects instead of controlled direct effects, because we wished to know what would happen if we intervened on neighborhood disadvantage (perhaps indirectly through housing voucher receipt, as in previous research (28)) but did not intervene on the school and peer environments. A second, practical reason for estimating these types of effects is that we can compare our estimates with those from the previous analysis (28).

Table 3. Risk Differences of the Effect of Living in a Disadvantaged Neighborhood on the Outcome (Adjusting for Covariates), Using Data From the National Comorbidity Survey Replication Adolescent Supplement, United States, 2001–2004

Outcome	RD	95% CI
Cigarette use	0.031	−0.014, 0.076
Alcohol use	−0.010	−0.059, 0.040
Problematic drinking	−0.016	−0.044, 0.012
Marijuana use	0.002	−0.042, 0.046
Problematic drug use	−0.007	−0.035, 0.022
DSM-IV diagnosis of substance use disorder	−0.009	−0.046, 0.029

Abbreviations: CI, confidence interval; DSM-IV, *Diagnostic and Statistical Manual of Mental Disorders*, *Fourth Edition*; RD, risk difference.

Total effects were estimated using a TMLE estimator (40).

We note that examining the association of a contextual exposure on an individual-level outcome, as we do here, does not directly affect our choice of estimand or estimator. However, variance estimation must account for clustering of individuals within neighborhoods, the options for which differ by type of estimator (e.g., bootstrapping, use of a sandwich estimator as we did for the Baron and Kenny approach, use of a sample-weighted influence curve as we did for TMLE, etc.). We refer the interested reader to Diez Roux (41) for a more in-depth discussion.

We imputed missing data using multiple imputation by chained equations (42), generating 30 imputed data sets. To fit models used in each estimator, we used the least absolute shrinkage and selection operator (lasso) (43, 44) This algorithm selects covariates to include in each model that improve model fit by more than 1 standard error from a high-dimensional list of main terms and 2-way interactions. The high-dimensional list of covariates included all potential measured confounders of the 1) $A - M$, 2) $A - Y$, and 3) $M - Y$ relationships and is given in Web Appendix 2.2. Age, race/ethnicity, household income, and sex were included in all models. We used 5-fold cross-validation to minimize the risk of overfitting. We used a sandwich estimator to calculate variance for each coefficient in the Baron and Kenny approach to account for sampling weights (45). We used the sample variance of the sample-weighted efficient influence curve to calculate variance for the TMLE approach. Last, we used a false discovery rate of 5% to account for multiple testing (46). R, version 3.3.1 (R Foundation for Statistical Computing), was used for all analyses.

Table 4. Risk Differences of the Effect of Living in a Disadvantaged Neighborhood on the Mediator (Adjusting for Covariates), Using Data From the National Comorbidity Survey Replication Adolescent Supplement, United States, 2001–2004

Mediator	TMLE		Baron and Kenny[a]	
	RD	95% CI	RD	95% CI
High violent crime at school	0.06	0.01, 0.12	0.07	−0.02, 0.16
Security at school	0.21	0.09, 0.34	0.27	0.07, 0.48
Most friends and siblings use marijuana	−0.01	−0.05, 0.04	0.00	−0.07, 0.08
No participation in sports or clubs	0.06	0.02, 0.11	0.04	−0.03, 0.11

Abbreviations: CI, confidence interval; RD, risk difference; TMLE, targeted minimum loss-based estimation.
[a] From Baron and Kenny (19).

Results

Illustrative example results are described in detail in Web Appendix 5. The analytical sample is described in Web Appendix 5.1 and shown in Web Table 1.

We first estimated the adjusted, marginal total effect of living in a disadvantaged neighborhood on each of the 6 outcomes considered (Table 3). Despite null total effects, pursuing causal

Table 5. Risk Differences of the Effect of Each Mediator on Each Outcome (Adjusting for Covariates), Using Data From the National Comorbidity Survey Replication Adolescent Supplement, United States, 2001–2004

Outcome-Mediator	TMLE		Baron and Kenny[a]	
	RD	95% CI	RD	95% CI
Cigarette use				
High violent crime	0.09	0.05, 0.13	0.07	−0.01, 0.14
School security	0.05	0.02, 0.08	0.02	−0.03, 0.06
No sports/clubs	0.10	0.06, 0.15	0.13	0.06, 0.20
Alcohol use				
High violent crime	0.06	0.03, 0.09	0.05	0.00, 0.10
School security	0.31	0.25, 0.37	0.29	0.21, 0.44
No sports/clubs	0.24	0.18, 0.29	0.18	0.11, 0.25
Problematic drinking				
High violent crime	0.15	0.11, 0.18	0.13	0.06, 0.20
School security	0.16	0.11, 0.21	0.23	0.11, 0.34
No sports/clubs	0.27	0.22, 0.32	0.29	0.19, 0.39
Marijuana use				
High violent crime	−0.02	−0.06, 0.01	−0.03	−0.10, 0.03
School security	−0.02	−0.04, 0.01	−0.04	−0.08, 0.01
No sports/clubs	−0.00	−0.04, 0.04	0.00	−0.06, 0.06
Problematic drug use				
High violent crime	−0.01	−0.03, 0.02	−0.01	−0.05, 0.04
School security	0.04	−0.00, 0.09	0.04	−0.04, 0.13
No sports/clubs	−0.01	−0.05, 0.04	0.02	−0.06, 0.10
Substance use disorder				
High violent crime	0.01	−0.02, 0.04	−0.01	−0.05, 0.02
School security	0.04	0.01, 0.08	0.07	0.00, 0.14
No sports/clubs	0.04	0.00, 0.08	0.07	−0.01, 0.14

Abbreviations: CI, confidence interval; RD, risk difference; TMLE, targeted minimum loss-based estimation.
[a] From Baron and Kenny (19).

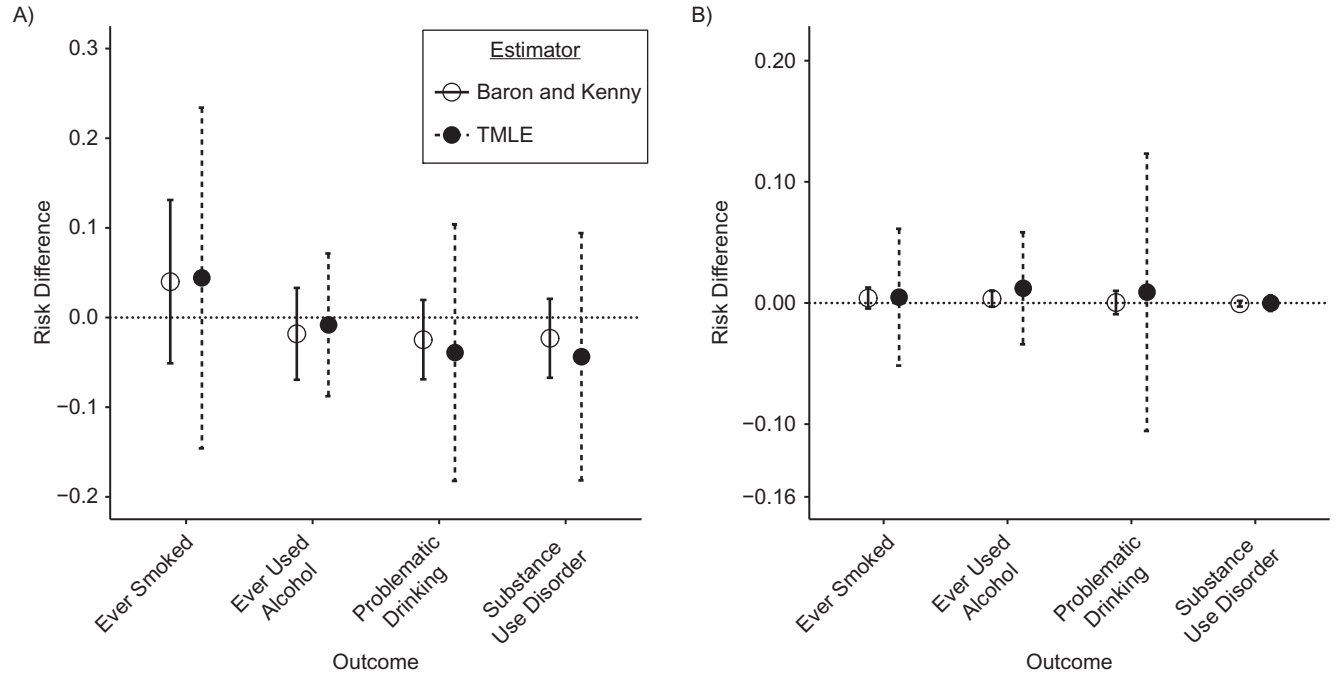

Figure 2. Direct (A) and indirect (B) effect estimates and 95% confidence intervals considering the mediator of high violent crime at school by outcome and mediation approach, using data from the National Comorbidity Survey Replication Adolescent Supplement, United States, 2001–2004.

mediation analysis might nonetheless identify important indirect effects that either offset each other and/or offset the direct effect (34).

We next estimated first-stage effects, which are the adjusted associations between exposure and each mediator (Table 4). For example, there is, on average, a 6% increased probability

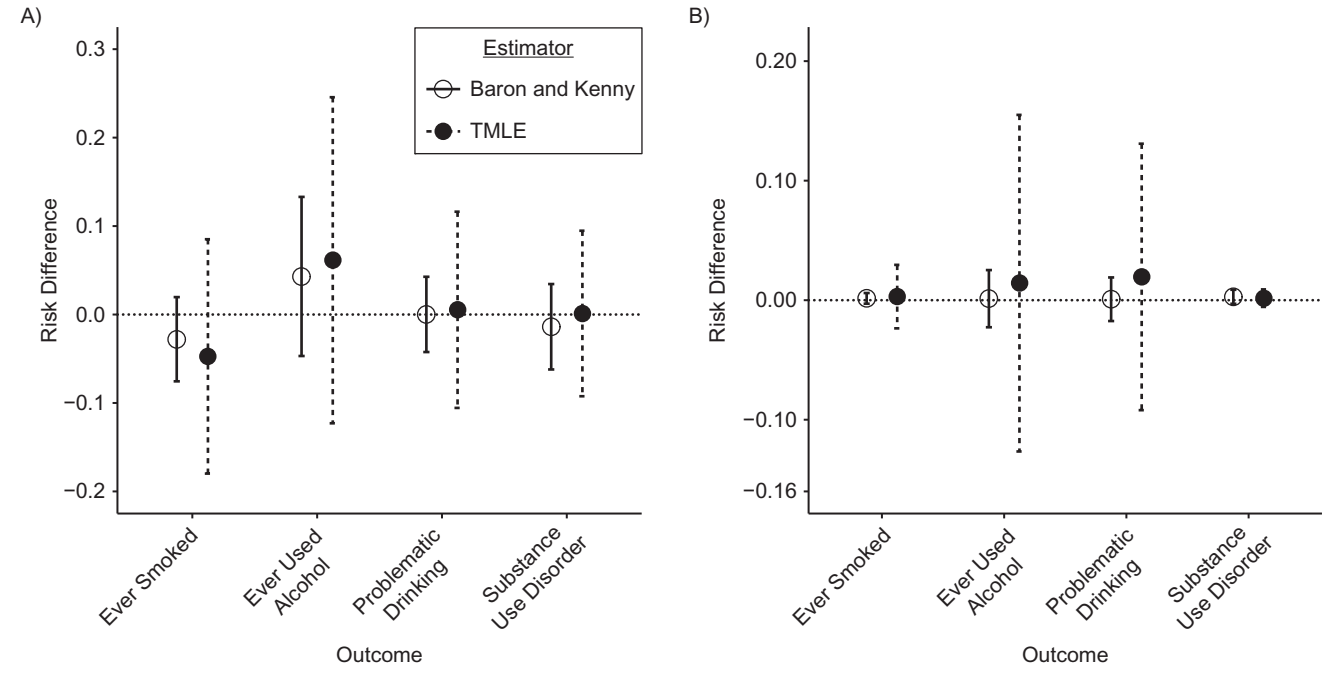

Figure 3. Direct (A) and indirect (B) effect estimates and 95% confidence intervals considering the mediator of security presence at school by outcome and mediation approach, using data from the National Comorbidity Survey Replication Adolescent Supplement, United States, 2001–2004.

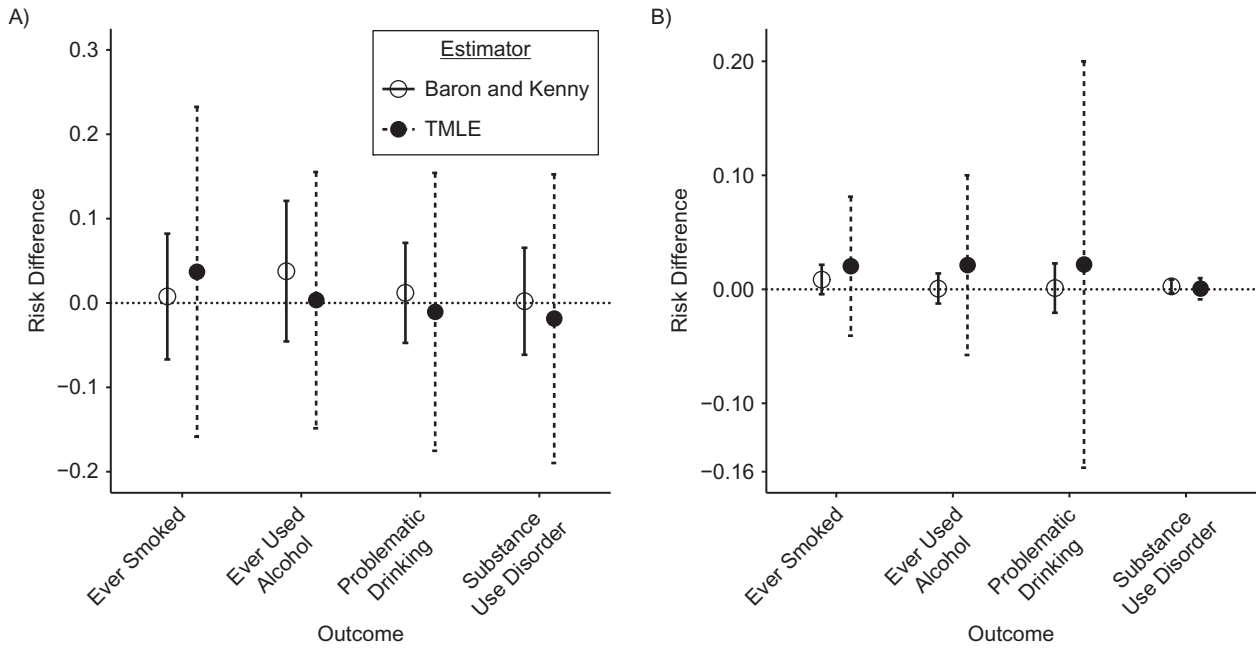

Figure 4. Direct (A) and indirect (B) effect estimates and 95% confidence intervals considering the mediator no participation in sports or clubs by outcome and mediation approach, using data from the National Comorbidity Survey Replication Adolescent Supplement, United States, 2001–2004.

(risk difference = 0.06, 95% CI: 0.01, 0.12) of attending a school with a high violent crime rate, under the scenario that everyone lives in a disadvantaged neighborhood versus everyone lives in a nondisadvantaged neighborhood. As seen in Table 4, living in a disadvantaged neighborhood is associated with attending a school with a high violent crime rate and a security presence as well as not participating in after-school sports or clubs. There is no association between living in a disadvantaged neighborhood and marijuana use by most peers and siblings. Point estimates of these first-stage effects are similar comparing the TMLE estimator with the regression estimator used in the Baron and Kenny approach. Confidence intervals are slightly wider using the Baron and Kenny approach and more often contain the null. Because there was no association between neighborhood disadvantage and most peers and siblings using marijuana using either estimator, we excluded this variable from subsequent analyses, given that it does not meet criteria for being a mediator (1).

We then estimated second-stage effects, which are the adjusted associations between each mediator and outcome (Table 5). For example, there is, on average, a 9% increased probability (0.09, 95% CI: 0.05, 0.13) of tobacco use, under the scenario that everyone attends a high violent crime school versus everyone attends a lower violent crime school. As seen in Table 5, there is no association between any of the mediators and the outcomes of marijuana use and problematic drug use using either estimator. Consequently, we excluded these outcomes from subsequent analyses. Point estimates of the second-stage effects are similar when comparing the TMLE estimator with the regression estimator used in the Baron and Kenny approach. Confidence intervals are slightly wider using the Baron and Kenny approach and more often contain the null.

Last, we estimated the direct and indirect effects of living in a disadvantaged neighborhood on each of the 4 remaining substance use outcomes through the 3 remaining mediators (Figures 2–4). All indirect effect estimates were null, even before implementing the more conservative false-discovery rate, so we cannot conclude that these aspects of the school and peer-based environment tested are on the pathway from neighborhood disadvantage to adolescent substance use.

Interpretation

We found similar results between the 2 estimators in terms of their point estimates (Figures 2–4). The Baron and Kenny approach resulted in narrower confidence intervals than the TMLE approach for the direct and indirect effect estimates. One reason could be that the indirect effect variance estimate using the Baron and Kenny approach is likely anticonservative, as discussed above and as compared with bootstrapped variance estimates (Web Figures 2–4). We compared the results of this analysis with a similar recent analysis conducted in the Moving to Opportunity program (28) (Web Table 1 and Web Appendix 5.2).

DISCUSSION

Understanding the mediation mechanisms by which an exposure or intervention affects an outcome can provide a look into what has been called a "black box" of many epidemiologic associations (47, 48), thereby providing further evidence of a relationship and possible points of intervention. Approaches for conducting mediation analyses have flourished recently, but formal instruction has generally not kept pace, leaving applied

researchers unsure of how to choose among the methods and how to implement newer, nonregression-based approaches.

In this work, we sought to address this confusion by providing an overview and tutorial of 1) mediation estimands and how to choose among them and 2) corresponding estimators and how to choose among them. However, this overview was limited in scope. We did not discuss mediation estimators that account for multiple mediators simultaneously (33, 37, 49–51) or estimators that incorporate sequential mediation (37). We also did not discuss nuances in adapting estimators for various exposure or outcome distributions (52–54).

We also provided a step-by-step illustration applying 2 approaches to examine mediation of the relationship between neighborhood deprivation and adolescent substance use by aspects of the school and peer environments in a national survey of mental health among urban adolescents. We compared the Baron and Kenny approach (18, 19), which is the most frequently used mediation estimator, with a recently developed TMLE substitution estimator of stochastic direct and indirect effects (16).

The Baron and Kenny approach is popular for its simplicity. However, this simplicity is at the expense of numerous potentially restrictive assumptions that might be at odds with both the data structure and research question. In addition, variance estimation under the Baron and Kenny approach might return inaccurate confidence intervals (3); evidence from bootstrapping suggested these confidence intervals in our illustrative example were anticonservative. The TMLE substitution estimation of stochastic direct and indirect effects requires fewer assumptions, is robust to model misspecification, and has theory-based variance estimation that results in appropriate confidence interval coverage even in finite samples (16). These advantages might come at the expense of simplicity. However, there is an R function that implements this estimator (16), and we include commented code in Web Appendix 4 for implementing both approaches.

In summary, mediation analysis is a key tool for understanding the mechanisms by which an exposure or intervention affects an outcome. In this work, we have presented an overview of the mediation approaches available and stepped through the decision-making process in choosing among them, with the goal of helping applied researchers identify and implement an approach that best aligns with their data structure and research question.

ACKNOWLEDGMENTS

Author affiliations: Department of Emergency Medicine, University of California, Davis, Sacramento, California (Kara E. Rudolph); Division of Epidemiology, University of California, Berkeley, Berkeley, California (Kara E. Rudolph, Dana E. Goin, Rebecca Crowder); Division of Genetic Epidemiology, National Institute of Mental Health, Bethesda, Maryland (Diana Paksarian, Kathleen R. Merikangas); Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland (Elizabeth A. Stuart); Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland (Elizabeth A. Stuart); and Department of Health Policy and Management, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland (Elizabeth A. Stuart).

This work was supported by the National Institute on Drug Abuse (grant R00DA042127; PI: K.E.R.) and the National Institute of Mental Health (grant R01MH099010; PI: E.A.S.). The National Comorbidity Survey Replication Adolescent Supplement and the larger program of related National Comorbidity Surveys are supported by the National Institute of Mental Health (grants U01-MH60220 and ZIA MH002808-11) and the National Institute of Drug Abuse (grants R01 DA016558) at the National Institutes of Health. The National Comorbidity Survey Replication Adolescent Supplement was carried out in conjunction with the World Health Organization World Mental Health Survey Initiative.

Conflict of interest: none declared.

REFERENCES

1. Valeri L, VanderWeele TJ. Mediation analysis allowing for exposure-mediator interactions and causal interpretation: theoretical assumptions and implementation with SAS and SPSS macros. *Psychol Methods*. 2013;18(2):137–150.
2. Petersen ML, Sinisi SE, van der Laan MJ. Estimation of direct causal effects. *Epidemiology*. 2006;17(3):276–284.
3. Shrout PE, Bolger N. Mediation in experimental and nonexperimental studies: new procedures and recommendations. *Psychol Methods*. 2002;7(4):422–445.
4. Imai K, Keele L, Tingley D. A general approach to causal mediation analysis. *Psychol Methods*. 2010;15(4):309–334.
5. Pearl J. Interpretation and identification of causal mediation. *Psychol Methods*. 2014;19(4):459–481.
6. VanderWeele TJ. Mediation analysis: a practitioner's guide. *Annu Rev Public Health*. 2016;37:17–32.
7. Lange T, Vansteelandt S, Bekaert M. A simple unified approach for estimating natural direct and indirect effects. *Am J Epidemiol*. 2012;176(3):190–195.
8. Naimi AI, Schnitzer ME, Moodie EE, et al. Mediation analysis for health disparities research. *Am J Epidemiol*. 2016;184(4):315–324.
9. Vansteelandt S. Estimating direct effects in cohort and case-control studies. *Epidemiology*. 2009;20(6):851–860.
10. VanderWeele TJ, Vansteelandt S, Robins JM. Effect decomposition in the presence of an exposure-induced mediator-outcome confounder. *Epidemiology*. 2014;25(2):300–306.
11. Merikangas K, Avenevoli S, Costello J, et al. National Comorbidity Survey Replication Adolescent Supplement (NCS-A): I. Background and measures. *J Am Acad Child Adolesc Psychiatry*. 2009;48(4):367–379.
12. Ogburn EL. Commentary on "Mediation analysis without sequential ignorability: Using baseline covariates interacted with random assignment as instrumental variables" by Dylan Small. *J Stat Res*. 2012;46(2):105–111.
13. Pearl J. Direct and indirect effects. In: Breese JS, Kolle D, eds. *Proceedings of the seventeenth conference on uncertainty in artificial intelligence*, San Francisco, CA: Morgan Kaufmann Publishers Inc; 2001: 411–420.
14. Didelez V, Dawid AP, Geneletti S. Direct and indirect effects of sequential treatments. In: *Proceedings of the Twenty-Second Conference on Uncertainty in Artificial Intelligence*, Arlington, VA: AUAI Press; 2006: 138–146.
15. VanderWeele TJ, Tchetgen Tchetgen EJ. Mediation analysis with time varying exposures and mediators. *J R Stat Soc Series B Stat Methodol*. 2017;79(3):917–938.

16. Rudolph KE, Sofrygin O, van der Laan MJ. Robust and flexible estimation of stochastic mediation effects: a proposed method and example in a randomized trial setting [published online ahead of print December 13, 2017]. *Epidemiol Methods*. (doi:10.1515/em-2017-0007).
17. Pearl J. *Causality: Models, Reasoning, and Inference*. 3rd ed. Cambridge, UK: Cambridge University Press; 2009.
18. Judd CM, Kenny DA. Process analysis: estimating mediation in treatment evaluations. *Eval Rev*. 1981;5(5):602–619.
19. Baron RM, Kenny DA. The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *J Pers Soc Psychol*. 1986;51(6):1173–1182.
20. Kling JR, Liebman JB, Katz LF. Experimental analysis of neighborhood effects. *Econometrica*. 2007;75(1):83–119.
21. Leventhal T, Brooks-Gunn J. New York City site findings: the early impacts of moving to opportunity on children and youth. In: Goering JM, Feins JD, eds. *Choosing a Better Life: Evaluating the Moving to Opportunity Social Experiment*. Washington, DC: The Urban Institute Press; 2003:213–244.
22. Theall KP, Sterk CE, Elifson KW. Perceived neighborhood fear and drug use among young adults. *Am J Health Behav*. 2009;33(4):353–365.
23. Leifheit KM, Parekh J, Matson PA, et al. Is the association between neighborhood drug prevalence and marijuana use independent of peer drug and alcohol norms? Results from a household survey of urban youth. *J Urban Health*. 2015;92(4): 773–783.
24. Tucker JS, Pollard MS, de la Haye K, et al. Neighborhood characteristics and the initiation of marijuana use and binge drinking. *Drug Alcohol Depend*. 2013;128(1–2):83–89.
25. Ennett ST, Flewelling RL, Lindrooth RC, et al. School and neighborhood characteristics associated with school rates of alcohol, cigarette, and marijuana use. *J Health Soc Behav*. 1997;38(1):55–71.
26. Zimmerman GM, Vasquez BE. Decomposing the peer effect on adolescent substance use: mediation, nonlinearity, and differential nonlinearity. *Criminology*. 2011;49(4): 1235–1273.
27. Bernburg JG, Thorlindsson T, Sigfusdottir ID. The neighborhood effects of disrupted family processes on adolescent substance use. *Soc Sci Med*. 2009;69(1):129–137.
28. Rudolph KE, Sofrygin O, Schmidt NM, et al. Mediation of neighborhood effects on adolescent substance use by the school and peer environments. *Epidemiology*. 2018;29(4): 590–598.
29. Kessler RC, Avenevoli S, Green J, et al. National Comorbidity Survey Replication Adolescent Supplement (NCS-A): III. Concordance of DSM-IV/CIDI diagnoses with clinical reassessments. *J Am Acad Child Adolesc Psychiatry*. 2009; 48(4):386–399.
30. Kessler RC, Avenevoli S, Costello EJ, et al. Design and field procedures in the US National Comorbidity Survey Replication Adolescent Supplement (NCS-A). *Int J Methods Psychiatr Res*. 2009;18(2):69–83.
31. Vansteelandt S, Daniel RM. Interventional effects for mediation analysis with multiple mediators. *Epidemiology*. 2017;28(2):258–265.
32. Rubin DB. Causal inference using potential outcomes: design, modeling, decisions. *J Am Stat Assoc*. 2005;100(469): 322–331.
33. Zheng C, Zhou XH. Causal mediation analysis in the multilevel intervention and multicomponent mediator case. *J R Stat Soc Series B Stat Methodol*. 2015;77(3):581–615.
34. Imai K, Keele L, Yamamoto T, et al. Identification, inference and sensitivity analysis for causal mediation effects. *Stat Sci*. 2010;25(1):51–71.
35. Robins JM, Greenland S. Identifiability and exchangeability for direct and indirect effects. *Epidemiology*. 1992;3(2): 143–155.
36. Zheng W, van der Laan M. Longitudinal mediation analysis with time-varying mediators and exposures, with application to survival outcomes. *J Causal Inference*. 2017;5(2):20160006.
37. VanderWeele T, Vansteelandt S. Mediation analysis with multiple mediators. *Epidemiol Methods*. 2014;2(1):95–115.
38. Tchetgen Tchetgen EJ. Inverse odds ratio-weighted estimation for causal mediation analysis. *Stat Med*. 2013;32(26): 4567–4580.
39. Zheng W, van der Laan MJ. Targeted maximum likelihood estimation of natural direct effects. *Int J Biostat*. 2012;8(1): 1–40.
40. van der Laan MJ, Rubin D. Targeted maximum likelihood learning. *Int J Biostat*. 2006;2(1).
41. Diez Roux AV. Estimating neighborhood health effects: the challenges of causal inference in a complex world. *Soc Sci Med*. 2004;58(10):1953–1960.
42. Buuren S, Groothuis-Oudshoorn K. Mice: multivariate imputation by chained equations in R. *J Stat Softw*. 2011;45(3).
43. Tibshirani R. Regression shrinkage and selection via the lasso. *J R Stat Soc Series B Methodol*. 1996;58(1):267–288.
44. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw*. 2010;33(1):1–22.
45. White H. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*. 1980;48(4):817–838.
46. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Series B Methodol*. 1995;57(1):289–300.
47. Greenland S, Gago-Dominguez M, Castelao JE. The value of risk-factor ("black-box") epidemiology. *Epidemiology*. 2004; 15(5):529–535.
48. Weed DL. Beyond black box epidemiology. *Am J Public Health*. 1998;88(1):12–14.
49. Steen J, Loeys T, Moerkerke B, et al. Flexible mediation analysis with multiple mediators. *Am J Epidemiol*. 2017; 186(2):184–193.
50. Daniel RM, De Stavola BL, Cousens SN, et al. Causal mediation analysis with multiple mediators. *Biometrics*. 2015; 71(1):1–14.
51. Nguyen QC, Osypuk TL, Schmidt NM, et al. Practical guidance for conducting mediation analysis with multiple mediators using inverse odds ratio weighting. *Am J Epidemiol*. 2015;181(5):349–356.
52. Lange T, Hansen JV. Direct and indirect effects in a survival context. *Epidemiology*. 2011;22(4):575–581.
53. Wang W, Zhang B. Assessing natural direct and indirect effects for a continuous exposure and a dichotomous outcome. *J Stat Theory Pract*. 2016;10(3):574–587.
54. Albert JM, Nelson S. Generalized causal mediation analysis. *Biometrics*. 2011;67(3):1028–1038.
55. VanderWeele TJ. Marginal structural models for the estimation of direct and indirect effects. *Epidemiology*. 2009;20(1):18–26.
56. Goetgeluk S, Vansteelandt S, Goetghebeur E. Estimation of controlled direct effects. *J R Stat Soc Series B Stat Methodol*. 2008;70(5):1049–1066.
57. Lendle SD, Schwab J, Petersen ML, et al. ltmle: an R package implementing targeted minimum loss-based estimation for longitudinal data. *J Stat Softw*. 2017;81(1):1–21.

58. Tchetgen Tchetgen EJ, VanderWeele TJ. On identification of natural direct effects when a confounder of the mediator is directly affected by exposure. *Epidemiology*. 2014;25(2):282–291.
59. Taguri M, Chiba Y. A principal stratification approach for evaluating natural direct and indirect effects in the presence of treatment-induced intermediate confounding. *Stat Med*. 2015; 34(1):131–144.
60. Vansteelandt S, VanderWeele TJ. Natural direct and indirect effects on the exposed: effect decomposition under weaker assumptions. *Biometrics*. 2012;68(4):1019–1027.